

# Phase Transition in the Number Partitioning Problem

Stephan Mertens\*

Universität Magdeburg, Institut für Theoretische Physik, Universitätsplatz 2, D-39106 Magdeburg, Germany  
(February 1, 2008)

Number partitioning is an NP-complete problem of combinatorial optimization. A statistical mechanics analysis reveals the existence of a phase transition that separates the easy from the hard to solve instances and that reflects the pseudo-polynomiality of number partitioning. The phase diagram and the value of the typical ground state energy are calculated.

64.60.Cn, 02.60.Pn, 02.70.Lq, 89.80.+h

Computer science has recently discovered the notion of phase transition in random combinatorial problems and its possible connections with algorithmic complexity. In such a context, statistical physics may provide an interesting perspective for understanding problems in theoretical computer science. In this letter we calculate the statistical mechanics of one of the core problems in theoretical computer science.

The number partitioning problem is an easily formulated optimization problem: Given a set  $\mathcal{A} = \{a_1, a_2, \dots, a_N\}$  of positive numbers, find a partition, i.e. a subset  $\mathcal{A}' \subset \mathcal{A}$ , such that the residue

$$E = \left| \sum_{a_j \in \mathcal{A}'} a_j - \sum_{a_j \notin \mathcal{A}'} a_j \right| \quad (1)$$

is minimized. A partition with  $E = 0$  is called *perfect*. The decision variant of the number partitioning problem is to determine if there is a perfect partition or not.

Number partitioning is of both theoretical and practical importance. It is one of Garey and Johnson's six basic NP-complete problems that lie at the heart of the theory of NP-completeness [1]. Among the many practical applications one finds multiprocessor scheduling and the minimization of VLSI circuit size and delay.

A partition can be encoded by numbers  $s_j = \pm 1$ :  $s_j = 1$  if  $a_j \in \mathcal{A}'$ ,  $s_j = -1$  otherwise. The cost function then reads

$$E = \left| \sum_{j=1}^N a_j s_j \right|, \quad (2)$$

and the minimum partition is equivalent to the ground state of the Hamiltonian

$$H = E^2 = \sum_{i,j=1}^N s_i a_i a_j s_j. \quad (3)$$

This is an infinite range Ising spin glass with Mattis-like, antiferromagnetic couplings  $J_{ij} = -a_i a_j$ . The thermodynamics of this model has been investigated by Fu [2] and recently by Ferreira and Fontanari [3].

Fu claims that in the random number partitioning problem "...no phase transition of any kind is found." [2]. If Fu were right, number partitioning would be a notable exception to the observation that many NP-complete problems do have a phase transition, parameterized by a control parameter that separates the easy from the hard to solve instances [4]. For random, integer  $a_i \in \{0, 1, 2, \dots, A\}$ , Gent and Walsh [5] proposed

$$\tilde{\kappa} = \frac{\log_2 A}{N} \quad (4)$$

as a control parameter: they found numerically that one typically has  $O(2^N)$  perfect partitions for  $\tilde{\kappa} < \kappa_c$ , whereas for  $\tilde{\kappa} > \kappa_c$  the number of perfect partitions drops to zero. The transition gets sharper with increasing  $N$ . Finite-size scaling lead Gent and Walsh to  $\kappa_c = 0.96$  for  $N \rightarrow \infty$ . This result contradicts Fu's claim, but as we will see now, this type of phase transition can indeed be found in the statistical mechanics of the number partitioning problem.

The canonical formalism of statistical mechanics requires the calculation of the partition function

$$\begin{aligned} Z &= \sum_{\{S_i\}} e^{-\frac{E}{T}} = \sum_{\{S_i\}} \int_{-\infty}^{\infty} dx e^{-|x|} \delta\left(x - \frac{1}{T} \sum_{j=1}^N a_j s_j\right) \\ &= 2^N \int_{-\infty}^{\infty} \frac{d\hat{x}}{2\pi} \prod_{j=1}^N \cos\left(\frac{a_j}{T} \hat{x}\right) \int_{-\infty}^{\infty} dx e^{-|x| + i\hat{x}x} \\ &= 2^N \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{dy}{\pi} \prod_{j=1}^N \cos\left(\frac{a_j}{T} \tan(y)\right) \end{aligned} \quad (5)$$

where  $T$  is the temperature. We write  $Z$  as

$$Z = 2^N \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{dy}{\pi} e^{N G(y)} \quad (6)$$

with

$$G(y) = \frac{1}{N} \sum_{j=1}^N \ln \cos\left(\frac{a_j}{T} \tan(y)\right). \quad (7)$$

\*email: Stephan.Mertens@Physik.Uni-Magdeburg.DE

At this point we could use the statistical independence of the  $a_j$  and replace the sum by the average of  $\ln \cos(\frac{a}{T} \tan(y))$  over  $a$ . This is usually done in Mattis-like spin glasses [6], but we will proceed without this substitution and calculate all thermodynamic quantities as functions of  $\{a_j\}$ .

For large  $N$  the integral in Eq. (6) can be evaluated using the saddlepoint technique. To find the saddlepoints of  $G(y)$ , we will assume that  $a$  can only take on values that are integer multiples of a fixed number  $\Delta a$ . For integer distributions  $\Delta a = 1$ , and for floatingpoint distributions  $\Delta a$  is the smallest number that can be represented with the available number of bits. This assumptions leads to an infinite number of saddlepoints, missed in [3],

$$y_k = \arctan\left(\frac{\pi T}{\Delta a} k\right) \quad k = 0, \pm 1, \pm 2, \dots \quad (8)$$

The resulting series of Gaussian integrals can be evaluated:

$$\begin{aligned} Z &= 2^N \sum_{k=0, \pm 1, \dots} \int_{-\infty}^{\infty} dy e^{-\frac{N}{2} G''(y_k) y^2} \\ &= 2^N \sqrt{\frac{2\Delta a^2}{\pi \sum_j a_j^2}} \coth \frac{\Delta a}{T}. \end{aligned} \quad (9)$$

From that we get the average energy

$$\frac{E}{T} = \frac{\Delta a}{T} \frac{\coth^2 \frac{\Delta a}{T} - 1}{\coth \frac{\Delta a}{T}} \quad (10)$$

and the entropy

$$S = N \ln 2 - \frac{1}{2} \ln \left( \frac{\pi \sum_j a_j^2}{2\Delta a^2} \right) + \tilde{S}\left(\frac{\Delta a}{2T}\right), \quad (11)$$

where the thermal contribution reads

$$\tilde{S}\left(\frac{\Delta a}{T}\right) = \ln \coth \frac{\Delta a}{T} + \frac{\Delta a}{T} \frac{\coth^2 \frac{\Delta a}{T} - 1}{\coth \frac{\Delta a}{T}}. \quad (12)$$

Note that for finite  $\Delta a$ ,  $\tilde{S}$  vanishes at zero temperature and increases monotonically with  $T$ . The entropy can be written as

$$S = N(\kappa_c - \kappa) \ln 2 + \tilde{S} \quad (13)$$

with

$$\kappa_c(N) = 1 - \frac{\ln\left(\frac{\pi}{6}N\right)}{N 2 \ln 2} \quad (14)$$

and

$$\kappa = \frac{\ln \frac{3}{\Delta a^2} \frac{1}{N} \sum_j a_j^2}{N 2 \ln 2}. \quad (15)$$

Note that  $\kappa = \tilde{\kappa} + O(\frac{1}{N\Delta a})$  for the distribution of the  $a_i$ 's considered by Gent and Walsh.

For  $\kappa < \kappa_c$  the entropy is extensive even for  $T = 0$ . According to Eq. (10), the corresponding energy is zero, hence we expect an *exponential number of perfect partitions*, in good agreement with the numerical results [5].

For  $\kappa > \kappa_c$  the zero temperature entropy seems to become negative. This would be wrong because the entropy must not be smaller than  $\ln 2$  for our discrete system. To see what is going on here, note that  $\kappa > \kappa_c$  means

$$2^{-N} > \Delta a \sqrt{\frac{2}{\pi \sum_j a_j^2}} \quad (16)$$

i.e. essentially  $\Delta a = O(2^{-N})$ . In this regime the contributions of  $\tilde{S}$  are  $O(N)$  for any finite  $T$ ,

$$\tilde{S}\left(\frac{\Delta a}{T}\right) = \ln\left(\frac{T}{\Delta a}\right) + 1 + O\left(\frac{\Delta a^2}{T^2}\right), \quad (17)$$

hence cannot be neglected. Technically we deal with this contribution by introducing an effective “zero” temperature  $T_0$  below which the system can not be cooled.  $T_0$  guarantees that the contribution of  $\tilde{S}$  remains  $O(N)$ . Its value can be calculated from the lower bound of  $S$ :

$$\begin{aligned} \ln 2 &= N(\kappa_c - \kappa) \ln 2 + \tilde{S}\left(\frac{\Delta a}{T_0}\right) \\ &\approx N(\kappa_c - \kappa) \ln 2 + \ln\left(\frac{T_0}{\Delta a}\right). \end{aligned}$$

From that we get

$$T_0 = 2\Delta a 2^{N(\kappa - \kappa_c)} = \sqrt{2\pi \sum_j a_j^2} 2^{-N}. \quad (18)$$

For  $\kappa > \kappa_c$  the ground state energy  $E_0$  reads

$$E_0 = T_0 = \sqrt{2\pi \sum_j a_j^2} 2^{-N} \quad (19)$$

This equation specifies the rigorous result that the median value of  $E_0$  is  $O(\sqrt{N} 2^{-N})$  [7].

To check Eq. (19) we consider the continuous variant of number partitioning, where the  $a_i$  are real numbers, uniformly distributed in the interval  $[0, 1)$ . In our formalism this means  $\Delta a \rightarrow 0$  and  $\sum_j a_j^2 = N/3$ . We are in the  $\kappa > \kappa_c$  regime and Eq. (19) becomes

$$E_0 = \sqrt{\frac{2}{3}} \pi N 2^{-N} = 1.447 \sqrt{N} 2^{-N} \quad (20)$$

In Fig. (1), Eq. (20) is compared to numerical data. The agreement is convincing. The prefactor  $\sqrt{\frac{2}{3}}\pi$  fits much better than the prefactor  $\sqrt{\frac{\pi}{6e^2}} = 0.2662$ , reported in [3].

To check whether  $\kappa(N)$  is a control parameter with a phase transition at  $\kappa_c(N)$ , we did numerical simulations. For fixed  $N$  and  $\kappa$  we calculated the fraction of instances that have at least one perfect partition. In accordance with Gent and Walsh [5] we find that this fraction is

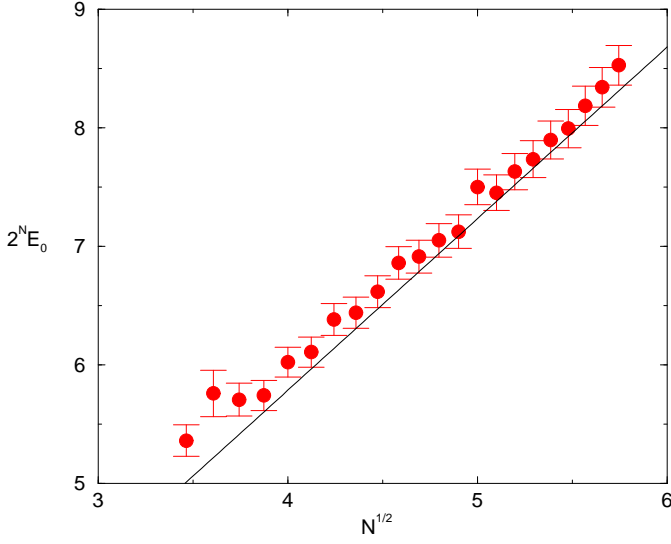


FIG. 1. Average minimum residue of the number partitioning problem with real numbers  $0 \leq a_i < 1$  compared to the analytical result Eq. (20) (straight line). Each data point is the average over  $10^4$  random samples.

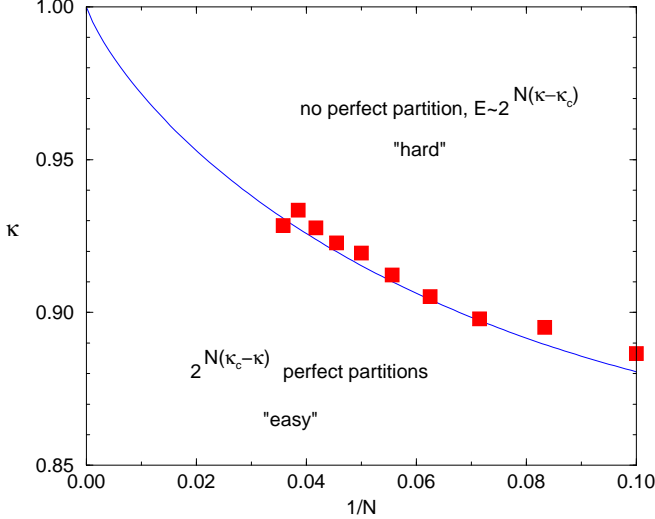


FIG. 2. Phase diagram of the random number partitioning problem.  $N\kappa$  is essentially the number of bits to encode the input numbers, see Eq. (15). The squares denote the phase boundary found numerically. The solid line is given by  $\kappa_c$  from Eq. (14). For  $\kappa < \kappa_c$ , the zero temperature entropy is extensive and a search algorithm typically finds quickly one of the  $O(2^N)$  perfect partitions. For  $\kappa > \kappa_c$ , no perfect partitions exist, and the optimization problem has a hard to find, unique solution.

1 for small  $\kappa$  and 0 for larger  $\kappa$ . The transition from 1 to 0 is sharp. Fig. (2) shows the numerically found transition points for  $10 \leq N \leq 28$  compared to  $\kappa_c(N)$  from Eq. (14). Again the agreement is convincing. Note that  $\kappa_c(N \rightarrow \infty) = 1$ . The asymptotic estimate 0.96 given by Gent and Walsh is probably due to the rather small values  $N \leq 30$  used in their simulations.

The two phases are very different with respect to the computational complexity of the corresponding instances. For  $\kappa < \kappa_c$ , a search algorithm is likely to find one of the numerous perfect partitions in short time, while the unique minimum partition for  $\kappa > \kappa_c$  requires exponential time to be found. This behavior can indeed be seen in numerical experiments [5,8,9]. Referring to the solvability of the typical instance we call the two phases “easy” and “hard”. In a numerical investigation, the precision of the numbers  $a_i$  is fixed and  $N$  is varied. For high precision and small values of  $N$ , we are in the “hard” phase: no perfect partition exists, and a search algorithm has to explore large parts of the configuration space. If one increases  $N$ , the search space growth exponentially, and so does the running time. On the other hand, increasing  $N$  gets us closer to the phase boundary. Beyond this threshold, the number of perfect partitions increases exponentially with  $N$ . A smart algorithm will try to find a perfect partition as early as possible. As can be seen from the experiments with the CKK algorithm [5,8,9], this may even lead to the effect that the running time now *decreases* with increasing  $N$ . As a function of  $N$  the running time has a sharp maximum at the phase boundary: The hardest problems are those close to the threshold. A similar behavior has been found in other NP-complete problems like, for example, the Satisfiability problem [10–12].

For bounded, integer values  $0 \leq a_i < A$  even the *worst case* complexity of number partitioning is polynomial in  $A$  and  $N$  [1]. This is no contradiction to the NP-completeness since a concise encoding of an instance requires  $N \log_2 A$  bits, and  $A$  is not bounded by a polynomial function of  $\log_2 A$ . Due to this property the number partitioning problem is called *pseudo polynomial* [1]. The exponential complexity of number partitioning relies on the fact that extremely large (or precise) input numbers  $a_i$  are allowed.<sup>1</sup> Pseudo polynomiality applies if the number of bits to represent  $a_i$  is fixed while  $N$  increases. This means, that  $\kappa$  decreases, i.e. we get into the “easy” phase for large enough  $N$ . Hence the notion “easy” not only refers to the *typical* (as shown here) but also to the *worst case* complexity. This is a notable feature of the number

<sup>1</sup> This distinguishes number partitioning from many other NP-complete problems like for example, the travelling salesman problem, which remains NP-hard even if the distances are restricted to take on the values 1 and 2.

partitioning problem: The statistical mechanics results hold beyond the typical case for which they are derived.

Looking at Eqs. (13, 19) an interpretation of the parameter  $\kappa_c$  suggests itself. Let the  $N$  numbers  $a_i$  each be represented by  $N\kappa$  bits. Now consider the residue  $E$  bitwise: About half of all partitions will set the most significant bit of  $E$  to zero. Among those partitions, about one half will set the second most significant bit to zero, too. Repeating this procedure we can set at most  $N\kappa_c(N)$  bits to zero until running out of available partitions. If  $\kappa < \kappa_c$ , we get a perfect partition before reaching this point. The remaining set of available partitions has  $2^{N(\kappa_c - \kappa)}$  elements. This explains the zero temperature entropy, Eq. (13). For  $\kappa > \kappa_c$ , the  $N(\kappa - \kappa_c)$  least significant bits in  $E$  can not be fixed by the optimum partition, leading to Eq. (19) for the residue.

## ACKNOWLEDGMENTS

The author appreciates stimulating discussions with Andreas Engel and Rémi Monasson.

- 
- [1] M. R. Garey and D. S. Johnson, *Computers and Intractability. A Guide to the Theory of NP-Completeness* (W.H. Freeman, New York, 1997).
  - [2] Y. Fu, in *Lectures in the Sciences of Complexity 1*, edited by D. L. Stein (Addison-Wesley Publishing Company, Reading, Massachusetts, 1989).
  - [3] F. Ferreira and J. Fontanari, J. Phys. A **31**, 3417 (1998).
  - [4] P. Cheeseman, B. Kanefsky, and W. M. Taylor, in *Proc. of IJCAI-91*, edited by J. Mylopoulos and R. Rediter (Morgan Kaufmann, San Mateo, CA, 1991), pp. 331–337.
  - [5] I. P. Gent and T. Walsh, in *Proc. of ECAI-96*, edited by W. Wahlster (John Wiley & Sons, New York, 1996), pp. 170–174.
  - [6] J. Provost and G. Vallee, Phys. Rev. Lett. **50**, 598 (1983).
  - [7] N. Karmakar, R. M. Karp, G. S. Lueker, and A. M. Odlyzko, J. Appl. Prob. **23**, 626 (1986).
  - [8] R. E. Korf, in *Proc. of the 14th IJCAI*, edited by C. S. Mellish (Morgan Kaufmann, San Mateo, CA, 1995), pp. 266–272.
  - [9] R. E. Korf, A Complete Anytime Algorithm for Number Partitioning, 1998, to appear in Artificial Intelligence.
  - [10] B. Hayes, American Scientist **85**, 108 (1997).
  - [11] R. Monasson and R. Zecchina, Phys. Rev. Lett. **76**, 3881 (1996).
  - [12] R. Monasson and R. Zecchina, Phys. Rev. E **56**, 1357 (1997).